DOCUMENT RESUME

ED 061 274 TM 001 209

AUTHOR	Melching, William H.
TITLE	Evaluation of Terminal Objectives in Reading.
INSTITUTION	Human Resources Research Organization, Fort Bliss, Tex. Div. 5.
PUB DATE	Apr 72
NOTE	14p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Achievement Rating; *Classification; *Educational Objectives; Formative Evaluation; *Opinions; Program Evaluation; Rating Scales; *Reading; Reading Comprehension; Reading Processes; Student Behavior; *Summative Evaluation; Task Performance; Verbs
IDENTIFIERS	*Terminal Objectives

ABSTRACT

This paper describes the results of an effort to apply a classification system, developed by the Human Resources Research Office (George Washington University), to a set of terminal objectives in reading. The classification system contained five factors, each factor embracing three levels. Some 312 terminal objectives (and 766 approximations) were classified by three raters. Agreement among raters was sketchy at the start of the rating, but improved dramatically with practice. It was concluded that (a) Classifying objectives is a useful procedure by which to evaluate the communicability of objectives; (b) Of all aspects of an objective, the verb is probably the most important; and (c) Reading objectives are no more difficult to classify than are objectives in other instructional content areas. (Author/CK)

ED 061274

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

EVALUATION OF TERMINAL OBJECTIVES IN READING

William H. Melching
HumRRO Division 5
El Paso, Texas 79931

For Presentation at the Annual Meeting of
The American Educational Research Association
Chicago, Illinois
April 3-7, 1972
ID #B-10; Session #3.3

TI 001 209

Abstract

This paper describes the results of an effort to apply a HumRRO-developed classification system to a set of terminal objectives in reading. The classification system contained 5 factors, each factor embracing 3 levels. Some 312 terminal objectives (and 766 approximations) were classified by 3 raters. Agreement among raters was sketchy at the start of the rating, but improved dramatically with practice. It was concluded that (a) Classifying objectives is a useful procedure by which to evaluate the communicability of objectives, (b) Of all aspects of an objective, the verb is probably the most important, and (c) Reading objectives are no more difficult to classify than are objectives in other instructional content areas.

AERA Presentation
Chicago, April 4, 1972
ID #B-10; Session #3.3

EVALUATION OF TERMINAL OBJECTIVES IN READING

William H. Melching
HumRRO Division 5
El Paso, Texas

A few years ago, a colleague and I[1] set out to clarify the methods, terms, and criteria associated with behavioral objectives. Based on our review of numerous sets of objectives prepared in various instructional contexts, we concluded that the most useful kind of objective was the terminal objective. We defined a terminal objective as a statement of the behavior expected of a student in a life or work performance situation. Additionally, we said that a terminal objective referred to behavior that is performed for its own sake. Thus, for example, while one might not copy letters of the alphabet in a life or work performance situation, he might construct words from such letters, especially if he arranged the words in a particular sequence.

Because we saw great variation in the level at which objectives were stated, we suggested a means by which terminal objectives might be meaningfully classified and/or evaluated. We offered five classification factors, each with three levels. The factors and levels are shown in Figure 1. With this classification scheme, we felt a trained reviewer could readily evaluate a set of objectives.

Recently, New York State Education Department (NYSED) adopted our concept of the terminal objective. NYSED undertook a full-scale effort to produce terminal objectives in reading and hired Independent Learning Systems (ILS)

---

[1] Ammerman, H. L. and Melching, W. H. The Derivation, Analysis and Classification of Instructional Objectives, Human Resources Research Office, The George Washington University, Technical Report 66-4, May 1966.

Classification of Terminal Objectives

Factor A: Type of Performance Unit	Factor B: Extent of Action Description	Factor C: Relevancy of Student Action	Factor D: Completeness of Structural Components	Factor E: Precision of Each Structural Component
Levels:	Levels:	Levels:	Levels:	Levels:
1. Specific Task	1. Fully Described	1. High Relevance	1. Fully Complete	1. Fully Precise
2. Generalized Skill	2. Partially Described	2. Moderate Relevance	2. Partially Complete	2. Partially Precise
3. Generalized Behavior	3. Stated Only	3. Low Relevance	3. Action Only	3. Vague

Figure 1

to develop the objectives. ILS was directed to develop terminal objectives consistent with the definition we had provided. In addition, NYSED requested that the developer produce two types of terminal objectives--a minimal set and a maximal set. The minimal set referred to reading behaviors that are required to maintain economic independence in a modern society, and to fulfill one's obligations as a parent and a citizen of the nation and community. By contrast, the maximal set referred to reading behaviors that are necessary to meet the reading requirements of tasks represented by the range of subject areas typically offered at the beginning of secondary school. In substance, then, two broad life-work performance situations were defined by NYSED.

A final requirement imposed on the developer by NYSED concerned the need to prepare a number of approximations (initially specified as four but subsequently reduced) for each terminal objective. An approximation was defined by NYSED as "an observable instance of behavior which is less complex, less difficult, requires more external support, or has a more limited range of manifestation relative to the terminal performance level. The approximation is further identifiable as the same form or type of behavior that is specified in the terminal performance objective."

To insure the production of acceptable objectives and to institute some

quality control procedures, NYSED engaged HumRRO to review and evaluate the objectives according to the classification scheme mentioned above. This, in turn, provided us an opportunity to explore the feasibility and usefulness of the classification system.

The remaining portion of this presentation discusses briefly some of our experiences in this regard. I have made no effort to sort experiences into categories, nor have I attempted to prepare a complete summary of the results of the classification effort. Instead I have elected to describe what we did, give a few results, and share with you some of the problems that we faced. To give you some idea of the scope of the review effort, I might mention that, excluding the objectives and approximations used in training sessions, a total of 312 terminal objectives and 766 approximations were evaluated by raters.

Some comments about the raters and their training would be an appropriate place to begin. Three psychologists at HumRRO Division 5 served as raters. Each was a Ph.D. and each had had some previous experience in preparing or evaluating objectives. None had had occasion to employ the Ammerman and Melching classification system, however.

Each rater was given a copy of the Ammerman and Melching report and a copy of a brief paper that elaborated on the definitions and descriptions of the levels of the five A&M factors. Once raters had studied these materials, they were directed to rate a sample of 11 terminal objectives. These objectives were constructed so that each level of each factor was a candidate rating in at least one instance. Each objective was rated independently by raters. They then met with the principal investigator, discussed the ratings, and attempted to reconcile differences. Following this, each rater was given 20

objectives prepared by ILS and asked to rate them independently. After this training, they were judged ready to perform their assigned functions.

The first set of objectives received for rating contained 87 objectives, all of the minimal set. Each objective was accompanied by four approximations. By and large, the objectives exhibited a fairly common feature--one that occasioned some distress in raters. This was the tendency for the objective writer to employ multiple verbs to describe the desired student action. While sample objectives had occasionally possessed this feature, this was not routine. Since the classification scheme assumes only one student action is stated, raters tried to handle the multiple verb objective by endeavoring to designate one verb as the prime one. In this regard they were not entirely successful.

Here's a somewhat exaggerated example of an objective with multiple verbs:

> Given a copy of the XYZ Buying Guide, the student will use the index to locate a specified product and will state if the product is considered acceptable or unacceptable and will correctly list at least one reason given in the Guide for this rating.

The inclination for writers to use several verbs when writing reading objectives may be peculiar to these kinds of objectives, but I doubt it. Let's look at the basic aspects of a reading objective. Presumably it requires the student to read a passage and then to behave in some way to evidence his comprehension of what he has read. So, for example, a student might be asked to read a label on a bottle of medicine and then indicate his understanding of the label by answering a question like, "What dosage should be given to a two-year old infant?" In other instances, an objective might require a student to perform two or more actions such as read, locate, copy down, restate, select, etc.

Writing objectives in this way, we believe, is due mostly to our previous writing habits. For convenience, we tend to serialize student actions. It saves space and covers a lot of ground. Unfortunately, such duality (or multiplicity) creates problems for us when we attempt to evaluate performance. Which action shall we assess? Which action is critical? How many different student behaviors are present?

Our solution to this multiplicity was this. Since actions of "read," "use," and "locate" are neither readily observable nor particularly informative, they might as well be deleted from statements. After all, if a student is given a written passage that asks him to perform in some way, it is assumed that he can't perform appropriately unless he has read it. What the writer must do when tempted to use multiple verbs is decide which student action is critical with respect to comprehension of the written material. So, for example, instead of asking a student to "use the telephone book to locate the number of the fire department and place a call," one might state the objective as:

> Given an emergency situation requiring the services of the fire department, the student must, with the aid of the local telephone book, indicate (state) the correct number of the department.

Whether he can successfully perform the actions of placing the call is an entirely different item of behavior.

Let's turn now to a related problem--related partly because it also concerns the use of verbs. This time, however, the focus is primarily on the relative specificity of the action verb. Again, perhaps because of earlier writing and speaking habits, the objective writer may employ verbs that are much too vague. I refer here to student action verbs such as "respond correctly, reply correctly, comply with, abide by," etc. Even "assemble correctly" can

be quite vague--a fact known by every parent who has attempted to assemble a child's toy according to instructions provided by the manufacturer.

The only solution to this problem, of course, is to require the writer to employ more specific verbs. And to do this, he must first determine explicitly just what the student is expected to do. As a result of this effort there is generally both an increase in student action specificity and a decrease in the tendency to use multiple verbs in a single objective.

With respect to the distribution of ratings on the first batch of objectives, Table 1 shows the results of the classification.[1/]

Table 1

Classification of 87 Terminal Objectives*

Level	Factor A	Factor B	Factor C	Factor D	Factor E
1	0	23	84	76	40
2	87	2	3	10	30
3	0	62	0	1	17

*The approximations were not classified according to the A&M system. Instead, they were subjected to another kind of evaluation, one not relevant to the main focus of this critique session.

The values in the table are concensus ratings. That is, if at least two raters agreed on a given level for a given factor of an objective, that was the rating which was recorded. When all raters disagreed, that is, when each rater selected a different level for a factor, it was necessary to invite raters to re-examine their classifications and to reconcile differences. These efforts encouraged exposure of unique interpretations and frequently promoted unanimity among judges.

[1/] To interpret the values in Table 1 and in subsequent tables, the reader should refer to the classification scheme shown in Figure 1, page 2.

At least four values stand out in the table. First, all objectives were rated as generalized tasks (Factor A, Level 2). This is what one might logically expect of reading behaviors. Second, student action was judged as "stated only" in 70% of the cases (Factor B, Level 3). Third, although raters experienced difficulty with verbs, they rated student action as highly relevant in almost every case (Factor C, Level 1). Thus, the raters agreed with the developer that the reading behaviors stated in the objectives were relevant to the designated work performance situation. Fourth, the vast majority of objectives were rated as fully complete with respect to structural components (Factor D, Level 1). In terms of preciseness of components, however, they were not so generous. They felt that over half of the objectives could be improved in precision.

All of the evaluations cited above, plus other more detailed feedback, were provided to the developer. In addition, of course, NYSED provided its own evaluation of the first batch of objectives to the developer. Armed with this information, the original objectives were scrapped, and a new attack on the preparation of objectives and approximations was undertaken. Within a short time, some 225 terminal objectives were produced by the developer; 61 were of the minimal set and 164 were of the maximal set. The number of approximations per objective now varied from two to four.[2/] A total of 766 approximations were prepared.

Before giving some of the results of this new classification effort, some comments may be made about the characteristics of the new objectives. With respect to use of verbs, the new objectives and approximations were notably superior to the original ones. Statements now embraced only one student action, and verbs describing highly specific action tended to be used. While

---

[2/] One objective actually had five approximations.

raters experienced little difficulty rating the minimal set objectives, this was not always the case for maximal set objectives. For example, raters insisted on revising one cluster of 18 maximal set objectives. Dealing with "relevancy" and "motivation," raters felt that, as written, the objectives assessed relevancy or attractiveness of the stimulus materials, not the capability of the student to read. In another cluster of 13 maximal set objectives where the student was required to "write in his own words a correct definition of a concept or thing," raters rejected the terminal objective and recommended that the first approximation be stated as an objective. Raters felt that the behavior required by the approximation was much more relevant to the intended work performance situation than the behavior required by the originally stated objective. A full statement of this approximation is as follows:

> Given a prose selection from a textbook that defines a concept or thing and a statement in which the concept of thing is used incorrectly, the student will state a corrected version of the statement.

Other results of the classification effort support the contention that the minimal set objectives were easier to rate than the maximal set objectives. Before presenting some of these results, however, it should be noted that this time raters classified both terminal objectives and approximations. Thus, 991 statements were classified. As before, if at least two raters agreed on the level of a factor, that agreement was accepted and recorded. Only in instances where there was maximal disagreement among raters (i.e., each rater selected a different level for a factor) was it necessary to have raters discuss their ratings and reconcile differences.

In Table 2 are shown the classification results for the minimal set objectives.

Table 2

Classification of 61 Minimal Set Terminal Objectives

		Factor				
		A	B	C	D	E
Level	1	0	0	61	38	23
	2	61	61	0	23	38
	3	0	0	0	0	0

The most striking thing about these values is that the ratings on the first three factors show no variation. All 61 statements were viewed by the raters as being generalized tasks, partially described, and highly relevant to an intended work situation. By contrast, raters felt the completeness and precision of structural components of many of the objectives were not fully acceptable. Now let's look at the classifications given to the maximal set terminal objectives. They appear in Table 3.

Table 3

Classification of 164 Maximal Set Terminal Objectives*

		Factor				
		A	B	C	D	E
Level	1	0	14	84	94	39
	2	149	144	56	64	119
	3	9	0	18	0	0

*Because 6 of the objectives were viewed as identical to a seventh objective, they were deleted and not rated.

With exception of the rating given to Factor C, the classification pattern of the maximal set objectives is not greatly different from that of the minimal set objectives. The difference in the intended work situation for maximal objectives versus minimal objectives probably accounts for the spread in relevancy seen in Table 3.

Tables 4 and 5 show how the approximations of the minimal and maximal sets of objectives were classified.

Table 4

Classification of 200 Minimal Set Approximations

Level	Factor A	Factor B	Factor C	Factor D	Factor E
1	0	1	131	135	99
2	200	199	47	65	101
3	0	0	22	0	0

The classification pattern of the minimal set approximations is more like that of the maximal set objectives than it is of the minimal set objectives. This is due primarily to the rating given to Factor C. Over 1/3 of the minimal set approximations were judged as being only moderate or low in relevance. Table 2 shows that none of the minimal set objectives obtained these ratings; all were judged highly relevant in student action.

Table 5

Classification of 566 Maximal Set Approximations*

Level	Factor A	Factor B	Factor C	Factor D	Factor E
1	0	0	159	358	239
2	492	477	318	170	289
3	36	51	51	0	0

*13 of the approximations were judged to be better terminal objectives than the objectives originally stated, and they were not included in this table. Also, 25 other approximations were not classified either because they were judged as incongruent with their associated objectives or because the objectives to which they belonged were deleted.

The classification pattern of the maximal set approximations appears to

be quite similar to the pattern of the maximal set objectives. Some minor variations can be found, but it is questionable that much meaning should be applied to them.

One final bit of information is interesting. This relates to the proportion of instances in which raters evidenced maximal disagreement in their judgments. A maximal disagreement was defined as any instance in which each rater assigned a different level for a given factor. The relevant information is shown in Table 6.

Table 6

Incidence of Maximal Disagreement Among Raters

Type of Statement	Number of Statements	Number of Disagreement Opportunities*	Actual Number of Disagreements	Number of Disagreements in Per Cent
Maximal Objectives First Group	116	580	68	11.7
Maximal Approximations First Group	403	2015	368	18.3
Minimal Objectives	61	305	0	0.0
Minimal Approximations	200	1000	8	.8
Maximal Objectives Last Group	48	240	1	.4
Maximal Approximations Last Group	160	800	2	.3

*Since each statement was rated on 5 factors, the number of opportunities for maximal disagreement was obtained by multiplying the number of statements in a group by 5.

Whether the first 116 objectives (and their 403 approximations) were infinitely more difficult than the remaining objectives, or whether there is simply a pronounced practice effect cannot be determined. But the drop in number of disagreements from start to finish is striking. Note too, that not a single maximal disagreement was recorded when raters were judging the minimal set objectives.

The following conclusions appear warranted:

1. Raters can learn to apply the A&M classification system, and they become more proficient in applying it with extended practice.

2. The act of classifying objectives is an effective means by which to examine the communicability of objectives.

3. The classification system provides useful information by which to evaluate objectives.

4. Of all aspects of an objective, the verb is probably the most important, especially with respect to judging the relevancy of the learner's behavior to the work performance situation.

5. Reading objectives are no more difficult to classify than are objectives in other instructional content areas.